# A Class of Fast Quaternion Valued Variable Stepsize Stochastic Gradient Learning Algorithms for Vector Sensor Processes

Mingxuan Wang, Clive Cheong Took, and Danilo P. Mandic

*Abstract*—We introduce a class of gradient adaptive stepsize algorithms for quaternion valued adaptive filtering based on three- and four-dimensional vector sensors. This equips the recently introduced quaternion least mean square (QLMS) algorithm with enhanced tracking ability and enables it to be more responsive to dynamically changing environments, while maintaining its desired characteristics of catering for large dynamical differences and coupling between signal components. For generality, the analysis is performed for the widely linear signal model, which by virtue of accounting for signal noncircularity, is optimal in the mean squared error (MSE) sense for both second order circular (proper) and noncircular (improper) processes. The widely linear QLMS (WL-QLMS) employing the proposed adaptive stepsize modifications is shown to provide enhanced performance for both synthetic and real world quaternion valued signals. Simulations include signals with drastically different component dynamics, such as four dimensional quaternion comprising three dimensional turbulent wind and air temperature for renewable energy applications.

## I. INTRODUCTION

QUATERNION valued signal modelling is undergoing fast development and is rapidly finding applications in statistical signal processing [1], [2], [3]. In the area of adaptive filtering, the recently introduced quaternion least mean square (QLMS) algorithm has offered a rigorous way to treat the real valued cost function of quaternion variables, and has formed a unified basis for adaptive filtering of three- and four-dimensional hypercomplex processes [4]. By virtue of the underlying division algebra, the QLMS caters naturally for the nonstationary dynamics and coupling between the components of three- and four-dimensional processes, and exhibits enhanced stability and more degrees of freedom in the control of the adaptation process as compared with its real-valued counterparts in $\mathbb{R}^3$ and $\mathbb{R}^4$.

The standard QLMS is second order optimal for processes with balanced power in the components of the signal model, that is, for second order circular (Q-proper) signals. This assumption is, however, not realistic for real world processes coming from vector sensor (3D ultrasonice anemometers, inertial bodysensors), where the signal components exbibit hugely different dynamics, and the measurements are subject to unisotropic noise. To make full use of the available second order information, the so-called augmented quaternion statistics has been recently introduced in the context of widely linear modelling [1]. This model uses both the information in the standard covariance and the three $(i, j, \kappa)$ pseudo-covariances, and exhibits enhanced stability and convergence

over its real valued counterparts in $\mathbb{R}^3$ and $\mathbb{R}^4$. The widely linear QLMS (WL-QLMS) uses such a model to enhance the accuracy and provide more degrees of freedom in the estimation of three-dimensional (pure quaternion) and four-dimensional (full quaternion) vector sensor signals.

The progress in sensor technology has also brought to light problems related to some of the mathematical deficiencies in multidimensional vector algebras. These include gimbal lock, where one degree of freedom is lost when e.g. two axes of the 3D gyroscope coincide. This can be explained by the fact that the product of two real vectors can be zero, even if none of the terms in the product is zero (non-division algebra). Division algebras, on the other hand, give a zero product only when one or both of the terms in the product are zero, thus providing an accurate model for fast rotation and orientation. The only division algebras are the fields of real $\mathbb{R}$, complex $\mathbb{C}$, quaternion $\mathbb{H}$, and octonion $\mathbb{O}$ numbers, and e.g. the benefits of widely linear adaptive filtering in the complex domain over filtering in $\mathbb{R}^2$ is well understood [5].

Following on our recent work on adaptive filtering in the quaternion domain [1][4], in this work we address the ways to make the learning rate parameter, which plays a prominent role in the performance in terms of convergence rate and steady-state misadjustment, adaptive. Fast and stable adaptation of filter coefficients is critical when operating in highly nonstationary environments, such as in the prediction of wind gusts and turbulences in renewable energy applications. To adapt the learning rate so that the algorithms best respond to the time-varying dynamics of the input signal, numerous gradient adaptive step-size stochastic gradient algorithms have been proposed both in the real and complex domain. These include Benveniste's algorithm, Farhang-Ang's algorithm, Mathews' algorithm, and the generalized normalized gradient descent algorithm [5].

For three- and four-dimensional vector processes, the separate component-wise processing in the real domain, or dual complex valued processing would be inadequate, despite a number of fast variable stepize algorithms that exist in these domains. To this end, in this work we introduce variable stepsize algorithms into the quaternion domain. The analysis is supported by simulations on both synthetic benchmark and real world quaternion valued processes, including 3D wind. It is also shown that the quaternion domain allows for a kind of heterogeneous data fusion, even when the dynamics of those data (such as in the joint modelling of air temperature and wind field) are radically different.

The authors are with the Department of Electrical Engineering, Imperial College London SW7 2AZ, UK (email: mingxuan.wang10@gmail.com, {c.cheong-took, d.mandic}@ic.ac.uk).

## II. FUNDAMENTALS OF QUATERNIONS

### A. Quaternion Algebra

The quaternion domain, a non-commutative extension of complex domain, provides a natural basis for the processing of three- and four-dimensional signals. A quaternion variable $q \in \mathbb{H}$ comprises a real or scalar part $\Re\{\cdot\}$ and a vector-part $\Im\{\cdot\}$, which includes three imaginary components, and $q$ can be expressed as:

$$
\begin{aligned}
q &= \Re\{q\} + \Im\{q\} = [q_a, \mathbf{q}] \\
&= \Re\{q\} + \Im_\imath\{q\}\imath + \Im_\jmath\{q\}\jmath + \Im_\kappa\{q\}\kappa \\
&= q_a + q_b\imath + q_c\jmath + q_d\kappa \qquad \{q_a, q_b, q_c, q_d \in \mathbb{R}\} \quad (1)
\end{aligned}
$$

The relationship between the orthogonal unit vectors, $\imath$, $\jmath$, $\kappa$, describing the three vector imaginary components of a quaternion is

$$
\begin{aligned}
\imath\jmath &= \kappa \quad \jmath\kappa = \imath \quad \kappa\imath = \jmath \\
\imath\jmath\kappa &= \imath^2 = \jmath^2 = \kappa^2 = -1 \quad (2)
\end{aligned}
$$

Observe that due to the noncommutativity of the quaternion product, we have $\jmath\imath = -\kappa \neq \imath\jmath$ for example. Other elements of quaternion algebra that are used in this work include the multiplication given by

$$
\begin{aligned}
q_1 q_2 &= [q_{a1}, \mathbf{q_1}][q_{a2}, \mathbf{q_2}] \\
&= [q_{a1}q_{a2} - \mathbf{q_1} \cdot \mathbf{q_2}, \ q_{a1}\mathbf{q_2} + q_{a2}\mathbf{q_1} + \mathbf{q_1} \times \mathbf{q_2}] \quad (3)
\end{aligned}
$$

where $q = q_a + q_b\imath + q_c\jmath + q_d\kappa$. Symbols "$\cdot$" and "$\times$" denote the dot-product and the cross-product respectively. The conjugate of a quaternion is given by $q^* = q_a - q_b\imath - q_c\jmath - q_d\kappa$ and the norm $||q||_2^2 = qq^*$. Note that quaternion conjugation is a self inverse mapping, and is termed an involution.

### B. Quaternion Involutions and the Widely Linear Model

Given a complex number $z = z_a + \imath z_b$, its real and imaginary part can be extracted as $z_a = \frac{1}{2}(z + z^*)$ and $z_b = \frac{1}{2\imath}(z - z^*)$. In the same context, the three perpendicular quaternion involutions given by [6]

$$
\begin{aligned}
q^\imath &= -\imath q\imath = q_a + q_b\imath - q_c\jmath - q_d\kappa \\
q^\jmath &= -\jmath q\jmath = q_a - q_b\imath + q_c\jmath - q_d\kappa \\
q^\kappa &= -\kappa q\kappa = q_a - q_b\imath - q_c\jmath + q_d\kappa \quad (4)
\end{aligned}
$$

can be exploited to calculate the four components of the quaternion variable $q$ as [6]

$$
q_a = \frac{1}{4}[q + q^\imath + q^\jmath + q^\kappa] \quad q_c = \frac{-\jmath}{4}[q - q^\imath + q^\jmath - q^\kappa]
$$
$$
q_b = \frac{-\imath}{4}[q + q^\imath - q^\jmath - q^\kappa] \quad q_d = \frac{-\kappa}{4}[q - q^\imath - q^\jmath + q^\kappa] \quad (5)
$$

Equation (5) thus demonstrates that any quadrivariate function $g(q_a, q_b, q_c, q_d)$ of real variables $q_a, q_b, q_c, q_d$ to be described as a function of the quaternion variable $q$ and all its $\perp$ involutions $q^\imath$, $q^\jmath$, $q^\kappa$. Similarly, the quaternion conjugate operation $(\cdot)^*$ can be expressed as a linear function of these $\perp$ involutions as

$$
q^* = \frac{1}{2}[q^\imath + q^\jmath + q^\kappa - q] \quad (6)
$$

In the context of mean square estimation (MSE), equation (5) suggests a convenient way to construct the corresponding widely linear model in $\mathbb{H}$ as

$$
y = \mathbf{g}^T\mathbf{x} + \mathbf{h}^T\mathbf{x}^\imath + \mathbf{u}^T\mathbf{x}^\jmath + \mathbf{v}^T\mathbf{x}^\kappa \quad (7)
$$

This model is by no means unique, as there exist different variants such as the one proposed in [2], however, it is adequate to capture the complete second order statistics in $\mathbb{H}$ [1].

## III. THE $\mathbb{H}\mathbb{R}$ AND $\mathbb{H}\mathbb{R}^*$ DERIVATIVES

A common optimisation objective in complex and quaternion-valued signal processing applications is to minimise the square error $\mathcal{J} = |e|^2 = ee^*$. This poses a problem, as the standard derivatives are not defined for functions dependent on variable $e$ and its conjugate $e^*$. To address this issue, Wirtinger considered the so-called pseudo-gradient,

$$
\nabla_\mathbf{w} J = \nabla_{\mathbf{w_r}} J + \jmath\nabla_{\mathbf{w_i}} J \quad (8)
$$

where $\mathbf{w}$ is a coefficient vector and subscripts 'r' and 'i' denote respectively the real and imaginary part of a complex variable and formalised it to what is known today as the $\mathbb{C}\mathbb{R}$ calculus. It was developed based on the isomorphism based on the fields $\mathbb{R}^2$ and $\mathbb{C}$ and the correspondence between complex-valued functions $f(z, z^*) = u(x, y) + \jmath v(x, y) \in \mathbb{C}$ and real bivariate functions $g(x, y) \in \mathbb{R}^2$ [5].

In the same spirit, we have recently extended the $\mathbb{C}\mathbb{R}$ calculus to the quaternion domain, termed the $\mathbb{H}\mathbb{R}$ calculus [6] to provide a unified framework for gradient based optimisation in $\mathbb{H}$. Due to space limitations, we summarise only the important results, for more details see [6]. The main component of the $\mathbb{H}\mathbb{R}$ calculus is given as

$$
\begin{aligned}
\frac{\partial f(q, q^\imath, q^\jmath, q^\kappa)}{\partial q} &= \frac{1}{4}\Big[\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_a} - \imath\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_b} \\
&- \jmath\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_c} - \kappa\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_d}\Big] \quad (9)
\end{aligned}
$$

whereas the $\mathbb{H}\mathbb{R}^*$ derivative can be summarised as

$$
\begin{aligned}
\frac{\partial f}{\partial q^*} &= \frac{1}{4}\Big[\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_a} + \imath\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_b} \\
&+ \jmath\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_c} + \kappa\frac{\partial f(q_a, q_b, q_c, q_d)}{\partial q_d}\Big] \quad (10)
\end{aligned}
$$

Similarly to the complex domain, it can be shown that the conjugate gradient $\nabla_{\mathbf{q}^*} f(\mathbf{q})$ gives the direction of the maximum rate of change of function $f(\mathbf{q})$ [6], making the $\mathbb{H}\mathbb{R}^*$ derivative in (10) a natural choice for gradient calculation and was considered in this work.

## IV. DERIVATION OF QUATERNION VARIABLE STEP-SIZE ALGORITHMS

We now formulate the update of the generalized normalized gradient descent (GNGD) algorithm [7], [8] in $\mathbb{H}$. For continuity, we first re-derive the QLMS algorithm [4] using the $\mathbb{H}\mathbb{R}^*$ calculus. Its cost function is given by $\mathcal{J}(k) = e(k)e^*(k)$, with filter output $y(k) = \mathbf{w}^T(k)\mathbf{x}(k)$, and its

weight update expressed as $\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}^*} \mathcal{J}(k)$, where

$$\nabla_{\mathbf{w}^*} \mathcal{J}(k) = e(k) \frac{\partial e^*(k)}{\partial \mathbf{w}^*(k)} + \frac{\partial e(k)}{\partial \mathbf{w}^*(k)} e^*(k) \quad (11)$$

Expanding the error expression, we have $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$ and $e^*(k) = d^*(k) - \mathbf{x}^H(k)\mathbf{w}^*(k)$. From the relationship (6), we have $\mathbf{w}(k) = \frac{1}{2}(\mathbf{w}^{\imath*} + \mathbf{w}^{\jmath*} + \mathbf{w}^{\kappa*} - \mathbf{w}^*)$; this permits the use of $\mathbb{HR}^*$ calculus, yielding $\partial \mathbf{w}(k)/\partial \mathbf{w}^*(k) = -1/2$. The error gradients thus become

$$\frac{\partial e^*(k)}{\partial \mathbf{w}^*(k)} = -\mathbf{x}^*(k) \qquad \frac{\partial e(k)}{\partial \mathbf{w}^*(k)} = \frac{1}{2}\mathbf{x}(k) \quad (12)$$

giving the QLMS update [4]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\Big(e(k)\mathbf{x}^*(k) - \frac{1}{2}\mathbf{x}(k)e^*(k)\Big) \quad (13)$$

Similarly, for the widely linear QLMS (WL-QLMS), the update can be expressed by [1]

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu\Big(\mathbf{x}^a(k)e^*(k) - \frac{1}{2}e(k)\mathbf{x}^{a*}(k)\Big) \quad (14)$$

where $\mathbf{w}^a(k) = [\mathbf{g}^T(k) \ \mathbf{h}^T(k) \ \mathbf{u}^T(k) \ \mathbf{v}^T(k)]$, and $\mathbf{x}^a(k) = [\mathbf{x}^T(k) \ \mathbf{x}^{\imath T}(k) \mathbf{x}^{\jmath T}(k) \mathbf{x}^{\kappa T}(k)]$.

### A. Variable Step Size Algorithms

A gradient adaptive learning rate $\mu(k)$ can be introduced into the QLMS algorithm (13), based on

$$\mu(k+1) = \mu(k) - \rho\nabla_\mu \mathcal{J}(k) \quad (15)$$

where the factor $\rho$ denotes the stepsize, and the gradient $\nabla_\mu \mathcal{J}(k)$ can be evaluated as

$$\begin{aligned}
\nabla_\mu \mathcal{J}(k) &= e(k)\frac{\partial e^*(k)}{\partial \mu(k-1)} + \frac{\partial e(k)}{\partial \mu(k-1)}e^*(k) \\
&= 2\Re\Big(e(k)\frac{\partial e^*(k)}{\partial \mu(k-1)}\Big) \\
&= -2\Re\Big(e(k)\mathbf{x}^H(k)\frac{\partial \mathbf{w}^*(k)}{\partial \mu(k-1)}\Big) \quad (16)
\end{aligned}$$

We have used the fact that

$$\frac{\partial e^*(k)}{\partial \mu(k-1)} = \Big(\frac{\partial e(k)}{\partial \mu(k-1)}\Big)^*$$

since the adaptive learning rate $\mu(k)$ is constrained in $\mathbb{R}$. Algorithms within this class can therefore be generalised as

$$\mu(k+1) = \mu(k) + 2\rho\Re(e(k)\mathbf{x}^H(k)\mathbf{\Phi}(k)) \quad (17)$$

where

$$\mathbf{\Phi}(k) = \boldsymbol{\beta} + \mathbf{x}(k-1)e^*(k-1) - \frac{1}{2}e(k-1)\mathbf{x}^*(k-1) \quad (18)$$

In the context of Benveniste's update [9], $\boldsymbol{\beta}$ can be evaluated as

$$\begin{aligned}
\boldsymbol{\beta} &= \mathbf{\Phi}(k-1) - \mu(k-1)\mathbf{x}(k-1)(\mathbf{x}^H(k-1)\mathbf{\Phi}(k-1)) \\
&\quad + \frac{1}{2}\mu(k-1)(\mathbf{\Phi}^H(k-1)\mathbf{x}(k-1))\mathbf{x}^*(k-1)
\end{aligned}$$

whereas for Farhang's algorithm [10], the parameter $\boldsymbol{\beta}$ becomes $\alpha\mathbf{\Phi}(k-1)$, whereas $\boldsymbol{\beta}$ vanishes for Mathews's

algorithm [11] in (18). As the updates of this class of variable stepsize algorithms and the Generalised Normalised Gradient Descent (GNGD) algorithm share similar terms, the derivation of this class of algorithms can be performed in the same way as that of GNGD, which is detailed in the next subsection.

### B. Generalised Normalised Gradient Descent Algorithm

The fixed variable step-size $\mu$ is replaced by an gradient adaptive stepsize

$$\eta(k) = \frac{\mu}{||\mathbf{x}(k)||_2^2 + \varepsilon(k)}$$

where $\varepsilon(k)$ is an adaptive regularization parameter. The stochastic gradient adaptation of the adaptive regularization parameter $\varepsilon$ can be performed as in (16), with $\partial \mathbf{w}^*(k)/\partial \mu(k-1)$ replaced with $\partial \mathbf{w}^*(k)/\partial \varepsilon(k-1)$, which can be obtained as

$$\begin{aligned}
\frac{\partial \mathbf{w}^*(k)}{\partial \varepsilon(k-1)} &= \frac{\partial \eta(k-1)}{\partial \varepsilon(k-1)}\Big(\mathbf{x}(k-1)e^*(k-1) \\
&\qquad\qquad -\frac{1}{2}e(k-1)\mathbf{x}^*(k-1)\Big) \\
&= -\frac{\mu\Big(\mathbf{x}(k-1)e^*(k-1) - \frac{1}{2}e(k-1)\mathbf{x}^*(k-1)\Big)}{\Big(||\mathbf{x}(k-1)||_2^2 + \varepsilon(k-1)\Big)^2} \quad (19)
\end{aligned}$$

Using (16)-(19), we can obtain the update of GNGD corresponding to the quaternion domain as shown in Table 1, together with Benveniste (BVSS) [9], Ang and Farhang (FVSS) [10] and Mathews and Xie (MVSS) [11] algorithms. These algorithms can be readily extended to widely linear QLMS (WL-QLMS) [12].

## V. SIMULATIONS

Two sets of simulations were conducted in a one-step ahead prediction setting in order to comprehensively assess the performances of the proposed quaternion adaptive step-size algorithms based on the widely linear (WL) model and the WL-QLMS with fixed learning rate [1].

The data sets used were

1) A stable Q-proper synthetic AR(4) process, with equal powers in each signal components, for which both the QLMS and WL-QLMS are theoretically optimal.
2) A Q-improper real-world 4D wind field signal, comprising three dimensional wind field as a pure quaternion with air temperature as a real part [4]. This signal has radically different dynamics and powers in signal components and is nonlinear and nonstationary.

The order of all the finite impulse response (FIR) filters was set to $N = 10$.

For both the synthetic and real world data used (see Figs. 1 and 2), the widely linear GNGD outperformed all the other algorithms in terms of both convergence rate and steady state performance. In general, the variable stepsize algorithms exhibited better performance than the fixed learning rate WL-QLMS, apart from Mathews' algorithm when processing the $Q-$improper wind field.

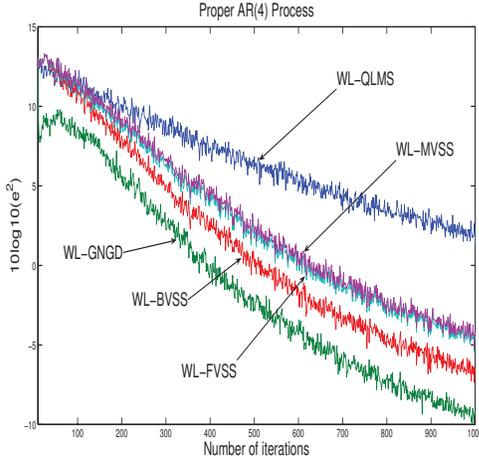| Algorithms | Update of the Learning rates |
|---|---|
| GNGD | $\eta(k) = \frac{\mu}{\|\mathbf{x}(k)\|_2^2 + \varepsilon(k)}$ |
| | $\varepsilon(k) = \varepsilon(k-1) - 2\rho\Re(\frac{\mu e(k)\mathbf{x}^H(k)(\mathbf{x}(k-1)e^*(k-1) - \frac{1}{2}e(k-1)\mathbf{x}^*(k-1))}{(\|\mathbf{x}(k-1)\|_2^2 + \varepsilon(k-1))^2})$ |
| Benveniste's VSS (BVSS) | $\mu(k+1) = \mu(k) + 2\rho\Re(e(k)\mathbf{x}^H(k)\mathbf{\Phi}(k))$ |
| | $\mathbf{\Phi}(k) = \mathbf{\Phi}(k-1) - \mu(k-1)\mathbf{x}(k-1)(\mathbf{x}^H(k-1)\mathbf{\Phi}(k-1)) + \frac{1}{2}\mu(k-1)(\mathbf{\Phi}^H(k-1)\mathbf{x}(k-1))\mathbf{x}^*(k-1) + \mathbf{x}(k-1)e^*(k-1) - \frac{1}{2}e(k-1)\mathbf{x}^*(k-1)$ |
| Farhang's VSS (FVSS) | $\mu(k+1) = \mu(k) + 2\rho\Re(e(k)\mathbf{x}^H(k)\mathbf{\Phi}(k))$ |
| | $\mathbf{\Phi}(k) = \alpha\mathbf{\Phi}(k-1) + \mathbf{x}(k-1)e^*(k-1) - \frac{1}{2}e(k-1)\mathbf{x}^*(k-1)$ |
| Mathews' VSS (MVSS) | $\mu(k+1) = \mu(k) + 2\rho\Re(e(k)\mathbf{x}^H(k)\mathbf{\Phi}(k))$ |
| | $\mathbf{\Phi}(k) = \mathbf{x}(k-1)e^*(k-1) - \frac{1}{2}e(k-1)\mathbf{x}^*(k-1)$ |



Fig. 1. Learning curves of the proposed variable stepsize algorithms on the prediction of the proper AR(4) process.
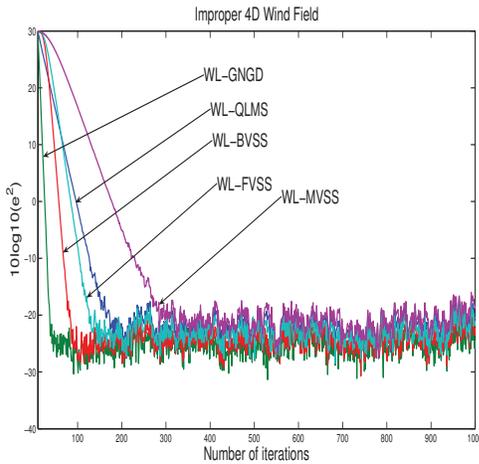


Fig. 2. Learning curves of proposed variable stepsize algorithms on the prediction of the improper 4D wind field.

## VI. CONCLUSIONS

We have proposed a novel class of variable stepsize algorithms for adaptive filtering of three- and four-dimensional processes modelled by quaternions. It has been shown that by virtue of its nonlinear stepsize update, the generalised normalised gradient descent (GNGD) algorithm provides the best performance of all the algorithms considered, as the other variable step size algorithms are based on 'linear' updates (which depend on estimators of $\partial\mathcal{J}/\partial\mu$). In other words the adaptive regularization term $\varepsilon$ within GNGD is a better choice for dealing with the different powers in signal components. However, Farhang's and Mathews' algorithms are computationally less involved than GNGD, and may be a preferrable choice when computational cost is critical.

## REFERENCES

[1] C. Cheong Took and D. P. Mandic, "Augmented Second-Order Statistics of Quaternion Signals," *Signal Process.*, vol. 91, no. 2, pp. 214 – 224, 2011.
[2] J. Vía, D. Ramírez, and I. Santamaría, "Properness and Widely Linear Processing of Quaternion Random Vectors," *IEEE Trans. on Information Theory*, vol. 56, no. 7, pp. 3502–3515, 2010.
[3] N. L. Bihan and J. Mars, "Singular value decomposition of quaternion matrices: A new tool for vector-sensor signal processing," *Signal Process.*, vol. 84, no. 7, pp. 1177–1199, 2004.
[4] C. Cheong Took and D. P. Mandic, "The Quaternion LMS Algorithm for Adaptive Filtering of Hypercomplex Real World Processes," *IEEE Trans. on Signal Processing*, vol. 57, no. 4, pp. 1316–1327, 2009.
[5] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, John Wiley and Sons Ltd, 2009.
[6] D. P. Mandic, C. Jahanchahi, and C. Cheong Took, "Quaternion Gradient Operator and Its Applications," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 47–50, 2011.
[7] D. P. Mandic, "A Generalized Normalized Gradient Descent Algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.
[8] S. C. Douglas, "Generalized gradient adaptive step sizes for stochastic gradient adaptive filters," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1396–1399, 1995.
[9] M. Metivier A. Benveniste and P. Prioutet, *Adaptive Algorithms and Stochastic Approximation*, Springer-Verlag: New York, 1990.
[10] W.-P Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size LMS algorithms," *IEEE Trans. on Signal Processing*, vol. 49, no. 4, pp. 805–810, 2001.
[11] V. J. Mathews and Z. Xie, "Stochastic gradient adaptive filters with gradient adaptive step size," *IEEE Trans. on Signal Processing*, vol. 41, no. 4, pp. 2075–2087, 1993.
[12] C. Cheong Took and D. P. Mandic, "A Quaternion Widely Linear Adaptive Filter," *IEEE Trans. on Signal Processing*, vol. 58, no. 8, pp. 4427 – 4431, 2010.